ANALYTIC PHILOSOPHY

Beware of Safety

Journal:	*Analytic Philosophy*
Manuscript ID	APHI-Dec-2017-OA-083.R2
Wiley - Manuscript type:	Original Article
Keywords:	safety, sensitvity, knowledge, epistemology, epistemic interest

Abstract: Safety, as discussed in contemporary epistemology, is a feature of true beliefs. Safe beliefs, when formed by the same method, remain true in close-by possible worlds. I argue that our beliefs being safely true serves no recognisable epistemic interest and, thus, that this notion of safety should play no role in epistemology. Epistemologists have been misled by failing to distinguish between a feature of beliefs — being safely true — and a feature of believers, namely being safe from error. The latter is central to our epistemic endeavours: we want to be able to get right answers, whatever they are, to questions of interest. I argue that we are sufficiently safe from error (in some relevant domain) by being sufficiently sensitive (to relevant distinctions).

Beware of Safety

ABSTRACT: Safety, as discussed in contemporary epistemology, is a feature of true beliefs. Safe beliefs, when formed by the same method, remain true in close-by possible worlds. I argue that our beliefs being safely true serves no recognisable epistemic interest and, thus, that this notion of safety should play no role in epistemology. Epistemologists have been misled by failing to distinguish between a feature of beliefs — being safely true — and a feature of believers, namely being safe from error. The latter is central to our epistemic endeavours: we want to be able to get right answers, whatever they are, to questions of interest. I argue that we are sufficiently safe from error (in some relevant domain) by being sufficiently sensitive (to relevant distinctions).

If you work on a building site, you'd better wear a helmet. Wearing a helmet keeps you safe from various kinds of injuries. Your head is not easily injured if you are thus protected. Other professions use other safety devices – whatever helps them to achieve their aims safely. Scientists, for example, conduct double blind experiments to exclude various distorting effects. Thus protected, they won't easily go wrong. Safety provides the achievement of an aim, be it the avoidance of injury or of error, with modal stability. Things won't go wrong easily if safety procedures are properly followed.

No wonder, then, that much of contemporary epistemology holds safety in high

regard. Sosa, Williamson, and Pritchard – to name three prominent epistemologists – have all endorsed safety in one way or another, and many others have followed them. On one such view, safety and knowledge go hand in hand. If a belief is unsafe, or true by luck, it is not knowledge. A safe belief, when true, is true with modal reassurance; it remains true in close-by possible worlds. The appeal of safety in contemporary epistemology has led to it replacing a different modal condition, sensitivity, which had been defended by Nozick, who argued that one can only know that p if, had p not been the case, one would not have believed that p.[1]

In this paper, I will argue that safety, as it is commonly understood by epistemologists, has no place in epistemology. I will criticize safety without participating in the production and discussion of ingenious counterexamples.[2] The point of these discussions is to question the necessity of safety for knowledge by appeal to our intuitions about various particular scenarios. My argument is different. I aim to show that a *concern* for safety would not be, in any recognizable sense, an *epistemic concern*. If this is correct, there is no reason for epistemology to be concerned with safety: with safety, that is, as it is commonly understood by epistemologists. I am not calling to put our trust in fate and, thus, forgo reasonable safety measures. Safety, in its normal

[1] Cf. Nozick (1981), 167-247.

[2] For example-focussed discussions of safety, see Brogardus (2013), Comesana (2005), Greco (2007), Kelp (2009), McEvoy (2009), Neta and Rohrbaugh (2003), Pritchard (2009) and (2012). In these discussions, Gettier's impact on the subject is acutely felt.

meaning, is important. Safe beliefs, however, are not.

The paper proceeds as follows. In section 1, after presenting standard definitions of safety, I explain why such a notion of safety has no place in epistemology. In section 2, I explain what has misled advocates of safety. They have disregarded the distinction between being safe from error on the one hand and safely believing a truth on the other. The former idea – being safe from error – is important. I will argue that one is safe from error if one is reasonably sensitive to whatever it is one wants to find out. I admit that the view proposed here – on which the only defensible understanding of being safe is to be sensitive – seems to go against much of what has been written on this topic in the last two decades. In section 3, however, I show that this view is, in fact, a natural view to hold, so that even advocates of traditional safety, on occasion, rely on it. Its spirit, if not its letter, it seems to me, has always been accepted.[3]

1 Why Safety Has No Place in Epistemology

1.1 Explaining Safety

I said that a belief that p is safe if it remains true in close-by possible worlds. In order to sharpen our idea of how safety is commonly understood and of how it differs from sensitivity, let us look at the ways in which its proponents have introduced safety.

[3] Please note that when I explain being safe from error in some domain by being sensitive to the relevant distinctions of this domain, I use a notion of sensitivity that, on occasion, will differ from Nozick's. I'm going to justify this use of the notion of sensitivity in section 2.

Sosa explains the difference between safety and sensitivity as follows.

> A belief is sensitive iff had it been false, S would not have held it […], whereas a belief is safe iff S would not have held it without it being true. For short: S's belief B(p) is sensitive iff not-p→not-B(p), whereas S's belief is safe iff B(p) →p (Sosa, 2000, 13 f).

The conditionals involved are subjunctives. Material conditionals allow contraposition, which would make safety and sensitivity logically equivalent; subjunctives don't contrapose.

The standard semantics for subjunctive conditionals uses the idea of possible worlds. We consider the world, closest to the actual world, in which the antecedent is true and ask whether the consequent is true in that world. If it is, the conditional is true, if it isn't, the conditional is false.[4] This semantic idea will be sufficient for our understanding of the sensitivity conditional. Our belief that p is sensitive, if we wouldn't believe that p, were p not the case. Like sensitivity, safety is meant to be an epistemologically significant property: not any true belief should count as safe. On the above semantics, however, safety would be satisfied by any true belief. In the closest world in which one believes that p, which in case of a true belief that p is the actual world, the belief that p is true. Thus, we need to extend the idea of how to evaluate

[4] See Stalnaker (1968) and Lewis (1973). For many intricacies concerning the semantics of conditionals, see Bennett (2003).

subjunctive conditionals to find a useful interpretation of safety. For the conditional to be true, the consequent should not only be true in the closest world in which the antecedent holds, it should continue to be true in a range of nearby possible worlds. This provides us with the understanding of safety outlined above: a true belief is safe if it continues to be true in a range of close-by possible worlds; the further away we have to look for a world in which our belief that p would be false, the safer the belief is.[5]

Safety for Sosa requires truth in all close-by possible worlds in which the belief is held on the same basis.[6] For Pritchard, safety is a necessary condition of knowledge. He

[5] In addition to sensitivity (and true belief), Nozick added a fourth condition to his analysis of knowledge, called 'adherence': p→B(p). Like advocates of safety, he needs to address the problem of explaining how a conditional with both true antecedent and true consequent can be a substantial condition on knowledge. Like advocates of safety, he strengthens the basic idea so that the consequent need not only be true in the closest but in a range of close-by worlds (in all of which the antecedent holds). Nozick offers the same semantics for his sensitivity and his adherence conditional. For further details, which will not matter here, see Nozick (1981, footnote 8, pp. 680-1). See Comesano (2007, 786f.) for a brief summary of Nozick's semantics for true-true subjunctives. The important point for my purposes here is this: on Nozick's account, sensitivity and adherence and, if we model safety accordingly, then also safety, all share the same semantics. I am grateful to a referee for this journal for asking me to clarify this matter.

[6] Since Nozick (1981, chapter 3.1) it is generally agreed that we need to restrict the comparison between the actual world and close-by possible worlds to cases in which the belief regarding p/not-p is held on the same basis, i.e. based on the same method. If, like in Nozick's case of the grandmother, the available method of epistemic access differs between a positive case – the grandchild being well – and a negative case – the grandchild being unwell – the fact that the grandmother is hindered from using the reliable method of looking at the child in the negative case, and is fed misleading testimonial evidence instead, does not show that the grandmother fails to know that the grandchild is

motivates the safety condition by the idea that safety precludes luck, as does knowledge. Pritchard offers the following definition of safety.

> An agent S has a safe belief in a true contingent proposition p $=_{df}$ in most near-by possible worlds in which S believes p, p is true. (Pritchard 2008, p. 446)

In some of his formulations the sameness-of-method idea I mentioned earlier is emphasized.

> S's belief is safe iff in most near-by possible worlds in which S continues to form her belief about the target proposition in the same way as in the actual world, and in all very close near-by possible worlds in which S continues to form her belief about the target proposition in the same way as [in] the actual world, her belief continues to be true. (Pritchard 2009, 34)

A succinct formulation that connects knowledge and safety is offered in Pritchard (2012, 253)

> If S knows that p then S's true belief that p could not have easily been false.

---

well in the positive case. This point about how Nozick wanted to understand the sensitivity requirement — namely as being restricted to the same method – has been adopted for safety by its proponents.

In our assessment of whether the belief that p remains true in close-by possible worlds, we restrict our attention to those worlds in which the belief that p is arrived at by the same method. For reasons that will not be important here (they have to do with continuous change), Williamson relaxes this condition. The method or, as he puts it, the basis of the belief, need not be the very same; it only needs to be relevantly similar.

> If in a case α one knows **p** on a basis *b*, then in any case close to α in which one believes a proposition **p*** close to **p** on a basis close to *b*, then **p*** is true (Williamson 2009a, 325).

Suppose the basis is actually the same. If one knows that p, p will remain true in all close-by worlds in which one believes p on the same basis: '… knowing 'on basis b' requires p to be true in all close worlds in which one believes p 'on basis b'' (Williamson 2009b, 21).

There are differences between the formulations provided by Sosa, Pritchard and Williamson. However, they clearly pursue a common idea. For a belief that p to be safe, it has to remain true in close-by possible worlds.[7]

1.2 What epistemic concerns are and how they figure in my argument

Having explained how safety is commonly understood in the relevant literature, I now

[7] This core idea -- a safe true belief remains true in most or all close-by worlds has -- is clearly present in all the current literature. See, for example, Vogel (2007, 83) who explains safety as follows: 'In all nearby possible worlds in which S believes P, P is true', or Blome-Tillmann (2009, 387)) who says, 'According to (SAFE), one only knows that p if one's belief p matches the facts in all nearby worlds.'

turn to questioning its importance for epistemology. I will argue that a concern for safety would not be an epistemic concern and, thus, it should be of no epistemological interest. This strategy requires two comments.

First, I need to explain what I mean by an epistemic concern. The paradigmatic epistemic concern is a concern for truth. We want true answers to our questions. If I am interested in what happened, this concern will be satisfied if what I believe about what happened is true. There might be other epistemic concerns, like believing in accordance with the evidence, reasoning along the lines of valid rules, or gaining insight, knowledge and understanding. I take no stance on the relation between such concerns and the concern for truth. As everybody will accept that a concern for truth is an epistemic concern, I can be liberal about what else belongs to this domain. Some things, however, won't. Besides epistemic concerns, we have prudential, aesthetic and moral concerns to name a few. A concern for happiness, for example, is not an epistemic concern. This is meant to leave it open whether all epistemic concerns are, in the end, based on other concerns, for example on prudential concerns. In order to be able to debate such issues, we need to make the distinction I am after. Wanting true answers to one's questions is an epistemic concern; wanting world peace and universal happiness is not an epistemic concern.

The second comment regards the methodology in play. I talk about concerns, about what we want, about where our interests lie, and I will conclude from the fact that an interest in safety would not be an epistemic interest that safety itself should play no

role in epistemology. This kind of argument is not standardly employed, so it needs explanation. In the explanation that follows I will start by assuming that knowledge requires safety.[8] (It will become clear that we can drop this assumption once my argument has been presented.)

There is what we can call a *de re* sense of wanting or of being interested in something. If you want to visit Austria then you want to visit the country of Hitler's birthplace even if it is unlikely that you want to visit Austria under this description. Similarly, if you want to marry the beautiful queen and she happens to be your mother, then you want to marry your mother. If you are not in a position to know that the queen is your mother, you are not in a position to know that you want to marry your mother. Nevertheless, wanting to marry the queen, who is your mother, makes it true that you

[8] This is the standard view taken by defenders of safety. Williamson emphasizes that claims about safety and intuitions about the closeness of worlds cannot provide an independent standard which could decide knowledge attributions. Safety, on his view, is a structural feature of knowledge; judgments about whether one knows are not based on judgments of safety; these judgements go hand in hand. '...someone with no idea of what knowledge is would be unable to determine whether safety obtained' (Williamson 2009, 305). Pritchard, in contrast, is involved in the post-Gettier project of looking for necessary and sufficient conditions for knowledge. Safety is necessary, though, as he now thinks, it is not sufficient for knowledge (Pritchard 2012). Sosa has endorsed safety as a condition for knowledge, for example in Sosa (2000), but since Sosa's (2007) development of his AAA theory, according to which knowledge is apt belief and apt belief is accurate because adroit, safety, on Sosa's view, has lost its standing as an integral part of knowledge. Sosa continues to have a positive attitude towards safety. For example, he thinks that the confusion between safety and sensitivity is responsible for the appeal of sceptical arguments. Pritchard combines traditional safety with Sosa's aptness view. He calls this combination 'anti-luck virtue epistemology'.

want to marry your mother in this sense of wanting. Consequently, if you are interested in knowledge, you are interested in whatever, according to the right theory of knowledge, turns out to be knowledge.

If I want my toast to be square, I want my toast to have sides of equal length. If I deny caring about the equal length of its sides, whilst insisting that it has to be square, I am confused. In order to illustrate a third possibility, in addition to conceptual confusion and not being in a positon to know, we need one more example. If I care about justice, and someone tells me that in order to act justly one has to maximize happiness, and, in some instances, what I care about it is contrary to what would actually maximize happiness, then I am committed to denying that in acting justly happiness is maximized. This denial might be evoked by considering particular scenarios. I certainly care about justice. Even if it would maximize happiness if the innocent were punished, I certainly object to the punishing of the innocent. So, I deny that justice would always require the maximization of happiness. It does not fit with my reflectively endorsed attitudes.

In general, if someone suggests that A is equivalent to B (or that A entails B) and I want A, but I deny wanting B, there are three options (illustrated by the three examples above). First, I am not in position to know that A and B are equivalent, secondly I am confused, and thirdly, I am committed to a rejection of the suggested equivalence (or implication). When it comes to suggestions about the nature of knowledge, the first option shall play no role: everyone is, when provided with the

right means of reflection, in a position to understand the nature of knowledge. If attitudes are not in line with suggested equivalences (or implications), we either have to deny these equivalences or we have to uncover some confusion that has led to attitudes which, given the facts, do not cohere.

1.3 A Concern for Truth and Two Kinds of Matching Concerns

At the base of our epistemic endeavours is a concern for truth. We want to find out what has happened or why or when and where. In general, we seek an answer to the question whether p or not-p.[9] If we get it right, there is a match between what we believe and what is the case. If we get it wrong, there is a corresponding mismatch and what we believe is not the case. At the base of our epistemic endeavours is a matching concern. This, I take it, is agreed on all sides.[10] What is not always recognized, however, is that there are different kinds of matching concerns.

Suppose you are in charge of folding the laundry and suppose only one kind of clothing was in the wash: socks. Your task then is to put all the socks into pairs. You have what we can call a symmetric or two-sided matching concern. The child's black sock needs a partner and so for all the others. Considering any pair, it does not matter

[9] Elsewhere, in Piller (2016) following Hieronymi (2005), I argue that in believing that p one answers (and takes oneself to have answered) the question whether p for oneself. This, I argue, explains why only p-related considerations can be reasons for believing that p.

[10] For the purpose of this discussion, I will disregard the idea that, in some situations, we may want to withhold and, thus, not want to commit epistemically at all, despite our awareness that one of the commitments would be correct.

where you start. You have the left sock and now you are looking for the matching right sock. Had you started with the right sock, you'd be now looking for the matching left sock. We encounter symmetry in this case because you are in control of fetching any sock you want. However, not all matching is such that one is in control of both sides. Think about how matching looks from the perspective of a particular sock. This sock, if it wanted to be helpful, would have a one-sided or asymmetric concern. If the sock could, it would express the following interest: 'I want to be put together with a sock that is similar to me'. The other option – whatever I am paired with, I want to be similar to the sock I am being paired with – is unavailable because no sock can change its size or colour.[11]

Take another matching concern. Hurrah for comfortable shoes! A comfortable shoe and the wearer's foot are nicely matched. They fit together in the right kind of way. The standard concern for comfortable shoes is an asymmetric, one-sided matching concern. The size of our feet is fixed and we look for a match by trying on different pairs of shoes. Only in the dark world of fairy tales do people cut off their heels and toes to make a shoe fit.

If we want to match one thing with another, our concern is symmetric if we want both: to find a match for the first kind of thing (left socks) and to find a match for the second kind of thing (right socks). Because we want both, it doesn't matter where we start.

[11] In order to sidestep an objection, we can imagine that toe-socks, which clearly distinguish between left and right, have become standard.

Our concern is asymmetric if we only want one kind of match. We want to match shoes to feet; it is not part of the standard concern for comfortable shoes that we wanted to match feet to shoes. (Though if we did, the shoes would be comfortable.)

1.4 The concern for truth is asymmetric

The basic epistemic concern, I said, is a matching concern. Hurrah for true beliefs! Is our concern for truth a symmetric or an asymmetric concern?

Like in the case of comfortable shoes, and unlike in the case of pairing socks, we usually have control over only one of the things we are supposed to match. Our option is either to believe p or to believe not-p. Like the size of our feet before, we now take the world as fixed and we try to produce a match by picking the right response: a belief that p if we are in a p-world and a belief that not-p if we are in a not-p-world. We try to create a match by picking the right belief. Think about the reversed concern which would be to match the world to our beliefs. Sometimes we can control some relevant aspect of the world. If I believe that someone in my department owns a Ford, I could ensure the truth of my belief by buying a Ford for myself (or for some colleague). From an epistemic perspective, this would be cheating. Buying a Ford, even though it is a fail-proof way to ensure the truth of this belief, is not an epistemic skill.

To manipulate the world to match what we take it to be is not an epistemic achievement. It does create a match, but it does not capture how standard epistemic

concerns operate: we take the world to be fixed and we want to have beliefs that match this fixed world. If this is so, the standard epistemic concern for truth is an asymmetric concern.

Suppose you believe that things will turn out for the best. Making them turn out for the best is driven by your desire for things going well which, according to the distinction we started with between a desire for truth and a desire for happiness, is not an epistemic concern. Likewise, the desire for having smaller (or larger) feet does not play a role when, on the basis of standard concerns, we buy new shoes.

Talking about desiring truth is, thus, ambiguous. Only one such desire has epistemic import. On the first reading of desiring truth, we want our beliefs to be true. What we want to be true, according to this reading, is a belief-state specified by its content. Suppose you believe that p. Ask yourself, do you want this belief to be true? The answer I'd give is that it really depends. If it is a good thing that p happens, I want p to happen and so I will also want my belief that p to be true. If, however, p is a bad or, as it might be, very bad, I do not want p to happen and so I also do not want my belief that p to be true. When I don't care whether p happens or not, I typically also don't care whether my belief that p turns out to be true. Suppose one thinks that more people will head towards the airport's exit than take the change-flight route. Does it matter at all whether on this occasion one will be right or wrong? Or, considering a case of wanting to be wrong, suppose you think that, unfortunately, the chances of a peace agreement between the fighting factions are minimal; does your epistemic

interest, i.e. your desire for truth, commit you in any way to wanting the fighting to continue? Why should truth-seekers want bad things to happen to them and others? Whether one wants one's belief that p to be true is, on this first reading, solely governed by one's non-epistemic concerns, i.e. by the value one assigns to p. This way of asking whether one wants one's beliefs to be true focusses on particular beliefs, i.e. it focuses on beliefs specified by their content. It is like asking whether you want a shoe with an already specified size to fit your foot whatever the size is. Our interest in comfortable shoes does not commit one to wanting a particular shoe of whatever size to fit. Our interest in truth does not commit one to wanting the world to be a certain way (so that it would match your belief). Suppose the shoe is of a size fit for a toddler. Do you have any reason which stems for your normal interest for comfortable shoes that the toddler's shoes are comfortable for you to wear? None, whatsoever! Do you have any reason which stems from your desire for truth to want the fighting to continue, if you believe that it will? None, whatsoever![12]

[12] Some might think, wrongly in my view, that in such a case one has a strong moral reason not to want the fighting to continue but that one would still have a weak reason, based on one's interest in truth, to want the fighting to continue. Note that, according to this view, there would be a conflict between one's epistemic and one's non-epistemic interest when we believe that a bad thing is going to happen. If our interest in truth is an asymmetric matching concern, the fact that we belief that p (like the fact that there are shoes of size S) provides us with no reason to want the world (or our feet) to be a certain way. Our ordinary desire for comfortable shoes does not commit one to wanting to have tiny feet. Our ordinary desire for truth does, similarly, not commit one to wanting anything bad to happen. We prefer a good world in which we believe that the bad is going to happen to a bad world in which our belief were true. This leaves the one-sided concern for a match intact, as it is still the case

There is, of course, a second way to understand the desire for truth. Do we want our beliefs to be responsive to the facts? In answering 'yes', we express what I take to be our central epistemic concern for truth. This differs from the first reading as we cannot specify the belief we want to have via its content. This is as it should be: a desire for truth is not wedded to believing that p; it is characterized by the openness of one's beliefs to whatever way the world may turn out to be. Whatever the world is like, may our beliefs match the world! This is the interest in truth that alone is relevant for epistemology.[13]

The interest in truth relevant for epistemology is an asymmetric matching interest. One wants to acquire beliefs that match the world. This interest does not commit one to care, on epistemic grounds, about matches that came about by changing the world. One wants to acquire beliefs that match the world, i.e. one wants to be open to the world and to be sensitive to the distinctions it contains. One need not want the world to match beliefs specified by their content. Such interests are not epistemic interests. We hold the world fixed in order to find out what we want to believe: in a p-world, we want to believe that p. We do not hold our beliefs fixed in order to hope for the

---

that for any mismatch there is match that is preferred. Whatever the world is like, we will want beliefs that match the world. This doesn't entail that whatever we believe, we want the world to match it.

[13] A dogmatist, in a colloquial sense, cannot stand to be wrong. He dismisses counterevidence as he hates to be proven wrong. Such a person has an interest that the world is as he believes it to be. In terms of virtue epistemology, such dogmatism exemplifies an epistemic vice.

world to match them. (And if we do so, what we want is guided by non-epistemic interests.)

1.5 Safety gets things the wrong way around

Now that we know how to understand an interest in truth, we can turn to attempts to strengthen our epistemic concerns modally. Remember what it is for a belief to be safe.

> An agent S has a safe belief in a true contingent proposition p $=_{df}$ in most near-by possible worlds in which S believes p, p is true. (Pritchard 2008, p. 446)

Suppose I wake up, get dressed and look at my watch: 'Oh my God, it is ten to nine!' Do I want this belief to be true? Certainly not: it means that I'll be late for my own wedding! I wish I were wrong in what I believe. I need to know. I shout out of the window to a passer-by. 'What's the time?' 'Ten to nine'! This is terrible. Do I have any epistemic reason to want it to be ten to nine as I am now certain it is? None, whatsoever! Do I want this belief to be safely true, i.e. true in those close-by worlds in which I hold the same belief having formed it on the same basis? Suppose I am wearing my grey socks but I could have worn my black socks as easily. If it weren't so late, I'd imagine myself standing there, now with black socks, believing it is ten to nine and realizing that I will be late for my wedding. Do I imagine myself wanting my belief to be true, or do I want from the perspective of the actual world that my belief be true in the world in which I am wearing black socks? No, I wish it would be at least

an hour earlier. My wedding is supposed to start at 9 o'clock.

If I believe that p, I have no epistemic reason, i.e. no reason that stems from a desire for truth properly understood, to want this belief to be true. As I have no reason to want it to be true in the actual world, I also have no reason to want it to be true in relevantly similar close-by possible worlds in which I hold this belief on the same basis. The fact that a belief is safe, i.e. that it remains true in close-by possible worlds, does not speak to any of my epistemic concerns. In the situation depicted above I have a strong epistemic concern. I really need to know what time it is. I have to get it right – it is very important. Wanting the belief I hold to be false does not interfere with the strong epistemic interest I display in the situation above. I engage in further enquiry by calling out to double check. I want to make sure that I believe that it is ten to nine if it really is and that I don't believe that it is ten to nine if it is any other time. I want to know the correct time, even if it hurts.

In this example one wants to know what time it is, and one wishes it were earlier than one thinks it is. One does not want the belief one holds to be true and so one does not want it to be safely true either. An interest in safety, may it be ten to nine in close-by possible worlds, would not be an epistemic interest. It is not related to an interest in truth, rightly understood, namely as an asymmetric matching concern. In our example an interest in safety would compete with what one wants most of all, which is not to

be late.[14]

The kind of argument I have offered should be familiar from other domains. I want justice. I don't want to punish the innocent, which, in the case at hand, would maximize happiness. So I deny that justice always requires the maximization of happiness. I want to know what time it is and I don't want the belief, which is, in the case at hand, that it is ten to nine, to be true or to be safely true. So I deny that wanting to know involves wanting one's belief to be true in the actual or in close-by possible worlds.[15]

[14] The father of the waiting bride might see his suspicions confirmed and mutter 'I told her so'. For him (but not for me), the situation holds something positive, at least he was right. This interest in being right, which commits him to wanting the world to be a certain way, is not a recognizable epistemic interest. I don't mean to deny that wanting to be right, i.e. wanting the world to confirm my beliefs, sometimes may serve epistemic purposes indirectly. If, for example, I have difficulties in understanding the display of my watch, the confirmation I receive from the passer-by is evidence that my ability to find out about the time is intact. No such concern about one's abilities need to be present. If there are no such concerns, there is nothing positive in seeing one's fears come true.

[15] Someone might try to turn the example around and say 'You don't want your belief that it is ten to nine to be safe – true; but you also don't want it to be knowledge because if it were knowledge it would be true. So safety and knowledge go hand in hand after all.' I want to know what time it is and, after having checked, I do know what time it is: it is 10 to nine. I do not want my belief that it is ten to nine to be knowledge. This is compatible with wanting to know rightly understood. It re-enforces the idea that the epistemic interest is an interest in believing that p, if p is the case and

My 'argument' points out something we knew along.[16] To be guided by truth is an interest about what one is like; it is an interest in one's own epistemic capacities; it is not an interest in what the world is like. Think back to Sosa's two conditionals. They have directionality, as they don't contrapose. As their directionality is opposed, it should be no surprise that one of them gets things the wrong way around.

Here is a summary of my argument. (1) A concern for truth is a one-sided matching concern. (2) Safety and sensitivity both deal with the match between beliefs and what is the case, i.e. between mind and world, but they differ in their directionality: safety goes from mind to world, whereas sensitivity goes from world to mind. Given (1), only one of them is of the right kind to capture our interest in truth, which is characterised by our openness to whatever the world is going to be like. In other words, the right concern (as well as its modal strengthening) must have world-to-mind conditionals as its content: I want, if the world is like this, to react in this way

---

believing that not-p, if not-p is the case. It is not an interest that attaches to a belief specified by its content. If I were to say that I want my belief that I won the lottery to be knowledge, this would be a roundabout way of saying that I'd wish I won the lottery. It wouldn't express an epistemic interest. Let me put this point in slightly different terms. Our epistemic interest is an interest in *whether* p is the case or, as in our example, *what* time it is. It is not an interest nor should it commit me to an interest in p on the grounds that I believe that p; it is also not an interest in my believing that p being a knowing of p. For a broader discussion of how epistemic and non-epistemic interests combine see Piller (2009).

[16] The argument offered does not commit me to any view about the nature of belief and how our interest in truth could explain this nature. If it did, I would be endorsing Anscombe's (1953) point – beliefs have a mind-to-world direction of fit – or Humberstone's (1992) claim that believing that p is an attitude characterised by the background intention not to believe that p if not-p. If Humberstone is right, a concern for a condition like sensitivity would be constitutive of believing.

and if the world is like that to react in that way. Thus, (3) a concern for safety does not relate to our concern for truth in the right way. It should play no role in epistemology.

2 Why This Point Has Been Missed and What Being Safe Amounts to

In the first section of this paper I have explained what epistemologists mean when they talk about safety. I have then argued that this notion of safety is of no epistemological interest. Once we understand this point – being concerned about truth in the right way does not mean that I want the beliefs I hold to be true or safely true – it should meet little resistance. My argument, I hope, brings something to light which everyone accepted all along. This attitude – what I've been arguing for should have been obvious all along – puts considerable argumentative burden on this section. If all is obvious, how could it have been missed?

In this part I offer an explanation of what went wrong in the epistemological debate in section 2.1. Safety was misconstrued as a property of beliefs when being safe is a property of believers. In section 2.2, I turn to a positive account of being safe. I argue that we are safe from error by being sensitive to relevant distinctions. In section 2.3, I show that safely believing, if it were an epistemic concern, would not suffice for being safe. My account of what being safe amounts to will often demand that one's believing satisfies a sensitivity condition. In section 2.4 I confront the challenge that sensitivity accounts have to be rejected because they violate closure, i.e. they violate the idea that single-premise inference from what is known would always expand our knowledge.

2.1 Being Safe from Error versus Safely Believing a Truth

The importance of safety measures in all sorts of domains is beyond dispute. They help to protect us from various harms we might otherwise suffer. What we want to achieve with the help of such measures is that *we* are as safe as we can be from harm. Being safe from harm is, if we follow this everyday conception of safety, primarily a property of persons. The epistemologist's conception of safety, in contrast, talks about the safety of beliefs. A belief that p is safe, we have said, if it remains true in close-by possible worlds.

Safety, on the everyday conception of safety, can be brought about by events that easily could have failed to happen. If lucky winds have blown me ashore, I am now safe from drowning. In terms of possible worlds, we say that, having reached the shore, there are only quite remote possible worlds in which I still drown. The winds could have easily blown in the opposite direction. But they have not. I am safe from drowning due to good fortune. The luck one has which makes one safe can be as big as it can get. Take winning the lottery. It keeps one safe from falling into poverty. Amongst those worlds in which one wins the lottery, there are only very few in which one wastes all one's fortune to become poorer than one was before. This is our first lesson to draw in this section: Being safe from injury or error is primarily a property of persons that ordinarily is brought about by contingent events. This contrasts with the epistemological notion of safety which is a property of beliefs.

Let us turn to sensitivity. Epistemologists think of sensitivity, as they did with safety,

as a property of beliefs. An agent's belief that p is sensitive, we have been told, if the agent would not have held this belief, were it not true.

Being sensitive, however, is better understood as a feature of persons. For example, I am sensitive to whether something is an insult or not. Even when dressed as a compliment, I can spot an insult straightaway. Being sensitive to the distinction between insults and things which are not insults means having the ability to categorize things correctly in terms of the distinction in question.

Consider another ability: the ability to fetch cold drinks from my fridge. The beer is cold so, when asked for a cold drink, I fetch a beer. However, had the beer been warm, I would have fetched a can of cold lemonade. It would sound odd if we said that it is a property of my fetching the beer that, had the beer been warm, it would have been a fetching of a cold lemonade. However, the conception of sensitivity used in epistemology offers the same (odd) picture. My believing that p is said to be sensitive if, had not-p been the case, it would have been a believing of not-p. But believings are ill described as turning into their opposite. It rather is I who, if not-p were the case, would hold a different belief. This is the second lesson I want to draw: sensitivity is also best understood as a feature of persons.

Epistemologists understand both safety and sensitivity as properties of beliefs. According to our ordinary conception of safety, in contrast, it is people who are, when things go well, safe from injury or error. Furthermore, it is people who are sensitive to distinctions. This suggests the idea that we are safe from error in some domain

when we are sensitive to the domain-specific distinctions.

2.2 Being Reasonably Safe by Being Reasonably Sensitive

The paradigmatic epistemic concern is a concern for truth. I have argued that such a concern is an asymmetric matching concern. Once we have a match, for example when we believe p in a p-world, this concern is satisfied. We might want more than simply to reach our aim (be it the avoidance of injury or of error): we might want to reach our aim safely, i.e. with the right kind of modal stability.

There is more than one way to understand such modal stability. Epistemologists offer various modal conditions as part of their quest for explaining the nature of knowledge. This is not my task here. My task is to explain how one can be reasonably safe in achieving one's epistemic aim by being reasonably sensitive to relevant distinctions.

I start with Nozick's understanding of sensitivity. According to Nozick, a person who correctly believes that p only knows that p if this person is sufficiently sensitive to the p/not-p distinction. In particular, had not-p been the case, one wouldn't know that p if one still believed that p in such a situation. Is one reasonably safe if one is sensitive in Nozick's sense? Suppose I truly believe that I am wearing black socks (I just had a look to confirm). If my powers of discrimination were weak, and I'd believe that I'm wearing black socks even if they were brown (or red), then, it seems, I'm not in a very strong epistemic position regarding the colour of my socks. In this example, worlds in

which I'd get it wrong are close-by (given my weak discriminatory capacity, I could have very easily worn socks of a different colour), so I'm not really safe in my getting it right on this occasion. Violating Nozick's sensitivity condition, would renders one's believing unsafe in this case.

It is well known that one's beliefs in the denial of sceptical hypotheses violate sensitivity: if one were indeed deceived by an evil demon, one would not think one was. This does not show, however, that one's epistemic positon in relation to sceptical hypotheses is as bad as the one of the person who can't distinguish socks of different colours. The difference between these cases (despite both believers being insensitive in Nozick's sense) is that, in the sock case, a world in which one fails to believe a truth is close by, whereas the world in which an evil demon deceives us is only a remote possibility. How safe we are, one could reasonably suggest, depends on the remoteness of worlds in which we get it wrong.

Note that even in the case of sceptical hypotheses, we find some modal stability in reaching the aim of believing correctly in cases when we, correctly, take ourselves not to be deceived. All close-by worlds will be such that the sceptical hypothesis is false and we believe truly in these worlds. This fact does not commit us to using the epistemologist's notion of safety. It is rather Nozick's notion of adherence, namely to believe p in close-by p-worlds, which accounts for this fact. In general, in strengthening the epistemic aim, we look at all close-by worlds, be they p-worlds or not-p worlds, and see whether in these worlds we believe appropriately. In this way

we respect the asymmetric nature of our concern for truth. This provides us with second way of understanding sensitivity: One is sensitive (and, thus, one would be safe in believing correctly), if, whatever the world is like, one would not easily get it wrong. The modal intuition that drives safety theories (not to get it wrong easily) is fully accounted for by this notion of sensitivity, which, typically, is a combination of Nozick's sensitivity and adherence conditions as long as there are close-by not-p worlds as well as p-worlds. If, however, there are no close-by not-p worlds, one can still be reasonably sensitive in virtue of the fact that one gets it right in all close-by worlds (all of which, in the case of denials of sceptical hypotheses, are p-worlds). Thus, there is a sense of being sensitive which is, in one way, weaker than Nozick's sense.[17]

There is a further, a third, way to understand being sensitive. It aligns being sensitive with a person's abilities. Ability ascriptions require reference to a set of conditions for

[17] This notion of being more or less sensitive is closely related to De Rose's notion of the strength of one's epistemic position. In the following summary, the two views are more or less the same. 'One's epistemic position with respect to P is stronger the more remote are the least remote possibilities wherein one's belief as to whether p does not match the fact of the matter (Sosa 1999, 144). In our context, it is important to insist that this account leaves it open what one believes in these possible worlds, i.e. it could be that one believes p or it could be that one believes not-p. (This is certainly in line with DeRose's position.) Thus this account is not about the stability of the relation between a particular belief and the world (as safety would have it), it rather concerns our responsiveness to the world whatever it may be like. See DeRose (1995) and the exchange between DeRose (2004) and Sosa (2004). DeRose (2004, 34) says that the best way to understand Sosa's safety makes safety very similar to his own notion of the strength of one's epistemic position. The important distinction between safety as a property of persons, i.e. being safe, and safety as a property of a specific belief does not come into clear enough focus. In trying to bring the two accounts in line, however, DeRose treats Sosa's safety as a property of persons. See DeRose (2004, 33) on what he calls S4-safety.

their normal manifestations. Usain Bolt has the ability to run very fast, but he can't run fast when submerged in water or when it is pitch dark. Such inability to run fast in some conditions does not undermine the ability we have in mind when we admire his running skills. Abilities are always abilities to perform in conditions which are normal for the exercise of the ability in question. The same, it seems reasonable to suggest, should hold for our epistemic abilities. I have the ability to distinguish red things from those that are not red. It does not undermine my self-ascription of this ability that there are borderline cases which I can't sort or that I fail in my sorting task when it is pitch dark or when a trickster make things that are not red appear red. One is reasonably sensitive to p/not-p if one manages to distinguish p from not-p under conditions which are normal for the exercise of the ability in question. What these normal conditions are will, obviously, differ for different instances of p/not-p.[18]

[18] This account of sensitivity has comparatively weak modal implications. If some success, epistemic or otherwise, is due to the exercise of an ability, I would still have succeeded, had the circumstances been different whilst still having been appropriate for the exercise of the ability. Such a weak modal account is compatible with Frankfurt's (1969) important anti-modal point which arose in his rejection of the idea that being responsible requires the ability to have acted otherwise. In Frankfurt's example we encounter an interfered with interferer and I agree with Frankfurt that prevented interference does not exclude responsibility. We are happy to assign responsibility in a Frankfurt case because what the agent did was an exercising of an ability. Although in the actual circumstances it would not have been possible to do anything different from what one did, had the conditions been different though normal, the agent would have still acted as he did. (I talk in more detail about the relevance of Frankfurt's point for epistemology in Piller, 2015). The idea I am endorsing here, namely the ability conception of being sensitive, goes back to Goldman (1979, 100), who says that '… the suitability of a belief-forming process is only a function of its success in 'natural' situations, not situations of the sort

The difference between the remoteness-of-mismatch account and the ability account of being sensitive comes to light when conditions which are unfavourable for the exercise of an ability are close-by. According to the latter account, one would still attribute sensitivity as an ability to the person in question, whereas on the former account, how sensitive one is will be affected by the closeness of these unfavourable conditions.

Epistemologists have mainly focussed on sensitivity as a candidate for a condition of knowledge. Thus, they were concerned about whether being insensitive, in the sense of violating Nozick's condition, would preclude knowing. I have advocated a switch of perspective: we should understand both safety and sensitivity as features of believers rather than as features of beliefs. Once we make this switch there is no obstacle to understanding being reasonably safe from error as being reasonably sensitive to relevant distinctions. I don't need to engage here with the question which level of sensitivity would be required for knowledge. What I do claim is the following: The degree in which we are safe from error varies with how sensitive we are. We

---

involving benevolent or malevolent demons, or any other such manipulative creatures.' There is another interesting issue associated with this account that I mention without pursuing it. Can we have a p-detecting ability without having the ability to detect not-p? I think we can. Sometimes this is simply a consequence of the specific instance of p. For example, I have a very good ability to detect that I exist but lack any ability to detect my non-existence. Thus, it is conceptually possible to have an ability to detect p without having any ability to detect not-p. We could have the ability to detect that we are awake, when we are, but no ability to detect that we are dreaming, when we are. For further discussion of this issue see Williams (1978, pp. 309-313), Humberstone (1988) and Williamson (1996).

might be more or less good in detecting some difference and, thus, more or less safe from error. How safe we want to be depends on the circumstances. We wear helmets on building sites but not full protective gear that could withstand enormous impacts because of the costs involved in ensuring such a high degree of safety. The very same considerations, I think, apply in the epistemic case. Our demands on how safe we want to be from error will depend on the practical costs of being wrong.[19]

Our interest in truth, I have argued, is asymmetric. Any modal strengthening of this aim will provide a conception of sensitivity. Sensitivity, in one form or another, is thus the only candidate of a modal component in one's epistemology. It alone strengthens

[19] In signal detection theory and in any kind of diagnostics (see, e.g., McNicol, 1972), we have to distinguish between two kinds of mistakes – false positives, believing that p in a not-p world, and false negatives, believing that not-p in a p-world. Depending on what is being investigated and for what purpose, false negatives and false positives will have different costs associated with them. Furthermore, the rarity of the condition investigated will be relevant. For a test for a very rare condition to be useful, the likelihood of false positives has to be very low – lower than the rarity. If the condition is common we need a low rate of false negatives. All this influences how to set the criterion which determines our response. It strikes me as significant that in signal detection theory, there is no need to consider whether a right response will remain accurate in those close-by worlds in which the response remains the same.

the basic epistemic aim of believing truly.[20]

2.3 Why the epistemologist's account of safety cannot explain what it is to be safe

In section 2.1 I have argued that we should distinguish between being safe and safely believing. I said that being safe is a desirable property of persons and not, like on the epistemologists's view of safety, a property of beliefs. I chose to focus on people to match our everyday conception of safety. This does allow that when a person is safe from error this safety-fact is explained by other facts, most notably one can be safe because of the method one uses in determining what is the case. In order to understand such safety – be it of a person or of the method the person employs – we have to refer to (some conception of) sensitivity. I have made this point in section 2.2. In this section, I complete my argument against safety as it has been discussed in epistemology. I will argue that safely believing something fails to explain why we are

[20] I have outlined three conceptions of sensitivity. I call them conceptions of sensitivity because they all strengthen the truth aim in the right way, i.e. they all respect the asymmetric nature of our concern for truth. I find the ability conception the most plausible. For the purposes of this paper, however, nothing depends on whether we choose the ability conception or whether our conception of sensitivity simply mirrors the domain of traditional safety by considering all or most close-by possible worlds. If we choose either of these conceptions, we will, under the name of sensitivity, endorse modal strengthenings of the truth aim that correspond to Nozick's adherence condition and not to his sensitivity condition which, for some not-p, might specify far-off possibilities or conditions that are not part of how we typically understand specific abilities. I hope this use will not give rise to any misunderstandings. It is justified on the grounds that, in contrast to safety, these conditions share with Nozick's original sensitivity condition the right kind of directionality which, if I am right, is an important aspect of our interest in truth.

safe from error when we are.

Remember that what makes one safe will usually be a contingent fact. Winning the lottery, I said, makes me reasonably safe from falling into poverty. If I win the lottery, I am safe from poverty due to a lot of luck. The same idea applies when we talk about beliefs. Suppose I wanted to know whether she loves me. I ask her and, suppose, she says 'No'. Assume furthermore that had she not won the lottery, she would have continued to pretend to love me and would have answered 'Of course, my dear'. But she did win and so she answered truthfully.[21] Now that she has told me truthfully, I know that she does not love me. I know because the interfering circumstance of her pretending to love me has been lifted, and when people are truthful or when they are bad pretenders, which she was not, I have the ability to detect their feelings. In this situation I am now safe from error due to the luck involved in a lottery win.[22]

Could one be safe from error in virtue of safely believing that p, i.e. can the epistemologist's conception of safety capture our ordinary idea of being safe? What would make it the case that my believing that p is safe, i.e. that it remains true in close-by possible worlds? The best-case scenario is one in which we find dual modal stability. The stability of continuing to believe that p is generated by assumption: we

[21] If we wanted a more cheerful case, we could turn the situation around so that only after winning the lottery was she able to confess her love for me.

[22] A structurally similar example, which struck me as rather far-fetched, appears in Neta and Rohrbaugh (2004).

are only considering the worlds in which one continues to hold the same belief. The modal stability of p would be ensured if p were not only true but true necessarily.

This best-case scenario could be put forward as a model of how to explain that we are safe from error in virtue of safely believing that p. We look at all the close-by worlds in which we believe that p and, in order to satisfy safety, we demand that p hold in all of them (which it will, if p is necessarily true). This has, however, the rather curious implication that we will always be safe when believing true necessities.[23]

From an intuitive point of view, this implication is curious because what makes me safe from error in, e.g., believing mathematical propositions, is not the modal status of mathematical truths but the fact that I am good at maths. Had I studied heraldry instead of times tables, I'd know lots about coats of arms and I'd be safe from error in that domain. The modal status of what we believe, it seems to me, has nothing to do with how safe we are. Instead of demanding dual rigidity in beliefs and what they are about, we should demand the right kind of flexibility in changing circumstances when we think about what it is to be good at something. [24]

The proximal explanation of the modal fact, being safe from error, invokes, on the

[23] Williamson's account of safety (see section 1) in terms of 'nearby propositions' believed on 'nearby bases' can be read as one kind of response to this problem.

[24] Pritchard, see the quotation in section 1, restricts safety to contingent propositions. This side-steps rather than confronts the issue. See Nozick (1986, 186f.) for a useful discussion of knowledge of necessities.

picture outlined above, yet another modal fact, namely to safely believe a truth. On the common-sense conception of being safe, by contrast, one is safe in virtue of a contingent fact.[25]

The idea that we can explain what it is to be safe from error via the idea of safely believing a truth encounters a further substantial problem. If one safely believes that p, p will hold in all those close-by worlds in which one continues to believe that p (on the same basis). We look at all the close-by worlds in which one believes that p, and if the belief that p is safe, p will hold in most or all those worlds. In order to assess the safety of a belief that p, we do not concern ourselves with worlds in which one does not believe that p. This, however, is a mistake.[26] There could be a close-by p-world in which one fails to believe that p and believes not-p instead. If such a possibility is close-by – believing not-p in a p-world – one is not safe from error despite one's belief that p being safe. Thus the notion of safety as used by epistemologists is not able to capture the idea of being safe from error. It neglects p-worlds in which one believes not-p and in such worlds one is in error.

Suppose whenever my brother believes that someone has insulted him he gets it right. This is due to the fact that he has a very high threshold of feeling insulted. One really

[25] It is possible to be safely safe. For example, I'd be safely safe from not being loved by anyone if God, who, if he exists, exists necessarily, cannot but love me. A modal fact – the love essential to the necessary being – makes me safe from not being loved by anyone and does so safely. This example, and its obvious rarity, shows that most of the time one's safety is brought about by contingent facts.

[26] DeRose (2004, 30) concurs.

needs to shout obscenities in his face in order for him to feel insulted. He is oblivious to any of the more subtle forms insults can take. Had the person not shouted, which, let us assume, could have easily been the case, he would not have recognised such behaviour as insulting. I advise him to take steps to raise his respective awareness. He should work on becoming more sensitive to what is and what is not an insult. Even though his insult-beliefs are safe, he still fall short of having an adequate epistemic ability.[27]

2.4 Hasn't Sensitivity Been Refuted?

In this paper I have argued for a negative thesis: safety has no place in epistemology. In support of the negative thesis I have offered an explanation of what went wrong in the debate: some epistemologists mistook safely believing, which is unimportant, for being safe from error, which is important. The latter notion provides a new home for the intuition which supported safety accounts, namely that we often want to achieve our aims with modal reassurance. I explained what is to be (more or less) safe from

[27] In terms of Nozick's account, the epistemologists' notion of safety does not capture his adherence condition which demands that in close-by p-worlds one believes that p which is substantially different from the safety demand that in close by believing-that-p worlds p holds. Williamson (2009b, 21) claims that when one knows that p, worlds in which one does not believe that p are not close-by worlds. I disagree. Suppose the angry man could easily have insulted my brother without shouting. Then he would not have believed that he had been insulted. (Assume furthermore that had the angry man not just won the lottery, which sufficiently increased his confidence, he would not have shouted.) The world in which he fails to shout and in which, due to my brother's high threshold, he fail to recognise this insult is, I'd think, a close-by world.

error via the idea that we are (more or less) sensitive to relevant distinctions: in order to be reasonably safe we have to be reasonably sensitive. The threshold of reasonableness depends, amongst other things, on the costs of getting it wrong in the two ways in which we may get things wrong (by either being too permissive or too restrictive in our methods). To complete the exposition of my view, I'd like to confront another idea which is popular within epistemology, namely that sensitivity accounts of knowledge have been refuted.[28]

As mentioned in the last section, my argumentative purposes do not require me to offer a theory of knowledge. My emphasis on sensitivity (in some form or another), however, make it of interest to explore the dismissal of sensitivity accounts of knowledge. This dismissal was brought about by two ideas. First, if we impose Nozick's sensitivity condition on knowledge, we cannot successfully reject sceptical hypotheses. Do we know that we are not brains in vats? If, in order to know, we need to be sensitive to the distinction between being a brain in a vat and not being one, we do not know that we are not brains in vats. If we were brains in vats, we would still believe that we are not brains in vats. (Assume that the closest world in which we are brains in vats leaves our experience unchanged.) Related to this point – sensitivity accounts cannot answer the sceptic – is another point: Sensitivity leads to a denial of epistemic closure, the principle which tells us that if one knows that p and knows that

[28] See, for example, Williamson (2000), Pritchard (2008), and Sosa (2000). For a contrasting view see Adams and Clarke (2005).

p entails q (and forms a belief about q on this basis) then one knows that q. On Nozick's sensitivity account, I know that I have hands (in the closest world in which I hadn't any, I'd notice), and I know (on conceptual grounds) that if I have hands I am not a brain in a vat. However, I do not know that I am not a brain in a vat because, as we have seen, this latter belief is not (sufficiently) sensitive.

I concentrate on the second of these points.[29] If we deny closure and claim that we need not know what we know is entailed by what we know, we seem, in a sense, to have abandoned logic. This strikes many as implausible.[30] If I know that my colleague Peter owns a Ford, I also seem to know that at least one of my colleagues owns a Ford. Logic, defenders of closure will say, helps us to enrich our knowledge.

Not all cases are alike. In the case of Ford ownership the evidence I have for Peter being the owner of a Ford is also evidence for the claim that one of my colleagues owns a Ford. However, evidence need not be transmitted along (known) logical relations. If I encounter an animal which looks like and smells like and behaves like a zebra, I have enough evidence to know that it is a zebra. I also know that zebras are not cleverly disguised mules. However, I have no evidence that the animal I

[29] In section 2.2 I have already argued that there are accounts of sensitivity which have exactly the same anti-sceptical force as safety theories. Both dismiss the sceptic on the ground that his scenarios are only remote possibilities. See De Rose (1995) for the development of such a view.

[30] For this reason Roush (2005), who defends a sensitivity account of knowledge, adds a closure principle. According to her view, one knows if one believes sensitively or if one can infer something from what one knows. As we will see, I am less sympathetic to the closure principle than she is.

encountered is not a mule made to look (and smell and behave) like a zebra. Thus, I don't know what is entailed by something I do know as long as such knowledge requires sufficient evidence.

I admit that my favourite sensitivity account, when offered as a theory of knowledge, would violate closure. I have the ability to recognise zebras by sight, i.e. I'm sufficiently sensitive to the most salient characteristics of zebras, namely their stripes. I don't have the ability to distinguish zebras from animals that have been turned into zebra look-a-likes. Thus, if I were to offer a theory of knowledge, I'd side with those who argue against closure. Empirical knowledge, they say, requires evidence. Evidence need not travel along the lines of logical entailment. Thus, knowledge is not closed under entailment. Knowledge, as they put it, is open.[31]

3. Appearances of the Right View

I have expressed the hope that the view advocated here, once clearly understood, will meet little resistance. This, I hope, is not simply an expression of my own dogmatism

[31] Sharon and Spectre (2017) make a strong case against closure and, in my view, they successfully undermine Hawthorne's (2004) case in favour of closure. Here is not the space to engage in any of the details of this debate. Let me only add that their case against closure differs from Nozick's, according to which the independent plausibility of sensitivity accounts is itself sufficient to embrace the consequence that closure has to be denied. Sharon and Spectre argue from the much weaker assumption that knowledge of empirical matters requires evidence. DeRose (1995, footnote 29) has claimed that the intuitive appeal of closure counts against Nozick's theory of knowledge. Sharon and Spectre have provided what DeRose has been asking for, namely an independent basis of denying closure.

when it comes to the rejection of traditional safety. Let me briefly look for indirect support of my optimism. I want to show how the view defended here shines through even in the work of advocates of traditional safety.

3.1 Timothy Williamson

Summarising what he did in earlier chapters, Williamson (2000, 147) writes, 'If one knows, one could not easily have been wrong in a similar case. In that sense, one's belief is safely true' (Williamson 2000, 147).

'If one knows *one* could not easily have been wrong.' Williamson focusses on what, according to the view presented here, is the central notion, being safe from error. 'In that sense, *one's belief* is safely true.' I have argued that in order to assess whether one is safe from error one needs to look at all close-by worlds, be they p-worlds or not-p worlds. If in most p-worlds one believes that p and in most not-p worlds one believes that not-p, then, I suppose, one will be reasonably safe from error. In order not to get things wrong easily, one needs a sensitive method (i.e. a relevant epistemic ability). The relevant notion of safety, being safe from error, will not be satisfied by any property of a belief, if this belief is already specified by its content.

'Safety and danger', Williamson stresses, 'are highly contingent and temporary matters' (Williamson 2000, 124). I have emphasized this point as well. In *Knowledge and Its Limits*, Williamson says that one can know something that does not obtain safely. 'One can believe that C obtains and be safe from error in doing so even if C

does not safely obtain, if whether one believes is sufficiently sensitive to whether C obtains' (Williamson 2000, 127). *Being safe* from error in regard to p/not-p requires *sensitivity* concerning p/not-p. At least in this quote, Williamson supports the view advocated here.

3.2 Ernest Sosa

Sosa (2004, 278) once wrote what I'd happily accept as a motto for this essay. 'Even once sensitivity and safety are distinguished, and even once we recognize that these are inequivalent contrapositives, it is still surprising just how different they are…' I agree: Only one of these conditionals speaks to our real interest in truth. Sosa continues, '… and how much more defensible safety is than sensitivity as a requirement for knowledge'. And here I disagree.

Sosa's views have developed. Aptness, success through adroitness or competence, has taken the place of safety in Sosa's virtue epistemology and aptness need not involve safety. 'What is required for a shot to be apt is that it is accurate because adroit, successful because competent. That it might easily have failed through reduced competence or degraded conditions renders it unsafe but not inapt' (Sosa 2007, 29).

What I say in this paper raises no objection to Sosa's virtue epistemology.[32] My

[32] I discuss Sosa's virtue epistemology in more detail in Piller (2015). It should also be noted that Sosa (2001, 50) says that wanting one's beliefs to be safe does not entail wanting them to be true. 'What we desire is only that our beliefs be safe; for any given proposition, other things equal we would generally desire this: that we would believe it only if it were true. Desire neither for the antecedent

picture adds the claim that epistemic competence is a matter of being sensitive to what the world is like but I don't think of this as a contentious claim.[33]

Finally I turn to two philosophers who have discussed worries that are related to mine. I want to show how my view differs from theirs.

3.3 Sherrilyn Roush and Duncan Pritchard

Sherrilyn Roush is, in a way, my closest ally. 'My main intuitive objection to safety as what decides whether a true belief is knowledge', Roush (2005, 121) writes, 'is that it gets the direction of fit wrong for what knowledge is'. Pritchard has also started to worry about direction-of-fit issues. Both provide examples about odd ways by which safety might be produced; they do not, like me, worry about the nature of safety and its relation to our epistemic concerns.

---

nor the consequent is logically entailed by our desire for the conditional. Our general antecedent desire is only for the safety of our beliefs, whatever they may be'. The mistake of safety theorists, I have argued, is to focus on a property of a belief specified by its content. In the quote above, Sosa senses the danger of any such view: If I want my belief that p to be safely true, I want my belief that p to be true; but this seems implausible: why should one want the bad to happen on the basis of thinking it is going to happen? 'If we believe a dear friend is terminally ill we would not want our belief to be true' (*ibid*). Despite his awareness of this danger, Sosa (2001) continues to provide his usual definition of when a belief that p is safe. In a different paper Sosa (2003) writes, 'More plausible seems the view that, for any arbitrary belief of ours, we would prefer that it be true rather than not true, other things being equal'. I have argued against this view.

[33] In section 2.3 I have argued that safely believing is compatible with being insufficiently sensitive. Here we learn that sensitive beliefs need not be safe. Contrast this view with Pritchard's who says that 'I am sympathetic to the idea that there is a way of thinking about the sensitivity principle such that it is equivalent to the safety principle (Pritchard 2012, fn 18).' I don't have such sympathies.

Roush's work is certainly in the neighbourhood of what I do. However, I am making a stronger point. She says that safety doesn't turn true belief into knowledge; I say that safety is of no epistemic significance. She says, that 'we want safe rather than unsafe beliefs' (Roush 2005, 122), that 'safety is a good property' (Roush 2005, 30), and that 'both sensitivity and safety are nice properties for a knowing belief to have' (Roush 2005, 123); I say we don't want our beliefs to be safe as, often, we don't want them to be true. The positive aspect Roush sees in being safe has, I suggest, wrongly been attributed by her to safe beliefs. Before I come back to her view, let me consider what Pritchard says about this issue.

Beliefs in propositions which are necessarily true could not have easily been false. As long as we hold the belief, e.g. that 2+2=4, it will always be true. Nevertheless, this fact – what I believe is necessarily true – does not seem to give my belief any privileged epistemic status. Instead of abandoning safety on these grounds, Pritchard points to what he takes to be a reformulation of safety. '[…] what we are interested in', Pritchard says, 'is rather how the agent forms her beliefs in similar circumstances and in response to the same stimuli. These beliefs may be beliefs that p, but equally they may be beliefs in distinct propositions' (Pritchard 2012, 257f). The mistake advocates of safety have made was to look for a feature of a belief specified by its content. Once we abandon this search and become interested in the ways a person believes, i.e. in the person's methods of belief-acquisition, whatever the specific belief may be a person holds, it seems most natural to be interested in beliefs that match the world, i.e. in

sensitive beliefs.[34]

In one of the examples that drive Pritchard's thinking, he comes close to the idea that safety gets things the wrong way round. Suppose a broken thermometer fluctuates randomly within a plausible range. An agent is in charge of providing temperature readings. His assignment is not permanent. Whenever he consults the thermometer, someone instantaneously adjusts the room temperature in accordance with these fluctuating readings. The agent forms a belief about the room temperature by consulting the broken thermometer.

Pritchard makes the following claims about this case. (1) The agent's beliefs about the room temperature exhibit the wrong direction of fit. (2) The agent, in this case, satisfies both safety and sensitivity. (3) The agent lacks knowledge in this case. (4) This case shows that no modal account can satisfy the idea that knowledge requires the exercise of an ability. In what follows I argue that, with the possible exception of (2), none of Pritchard's claims is plausible.[35]

[34] In a footnote Pritchard explains further. 'Although I have opted for the safety principle over the sensitivity principle here as the best way of thinking about the anti-luck intuition, I am sympathetic to the idea that there is a way of thinking about the sensitivity principle such that it is equivalent to the safety principle, although it should be noted that the sensitivity principle, so conceived, would be a very different beast to that put forward by folk such as Nozick' (Pritchard 2012, fn 18, 257). Pritchard, in this quote, is sympathetic to the idea that we are safe from error by using sensitive methods. The modal aspect of sensitivity can, as I have described in section 2.2 above, be understood in ways that indeed differ from Nozick's understanding.

[35] In his own words Pritchard 2012, 260f): (1) '… what is wrong with Temp's beliefs is that they exhibit

A thermostat regulates the room temperature. A thermometer measures it. We can say that a thermometer gives us a temperature reading with an assertion sign, a thermostat gives us a temperature reading with a command sign. What are taken to be asserted temperature readings are, actually, temperature commands which are, in the world of his example, immediately and successfully enacted. Thermostat readings and thermometer readings have different directions of fit. To use a standard reading of this difference, we can say that the world, i.e. the whole thermostat set-up, is to blame if there is a mismatch between the thermostat reading and the temperature, whereas the thermometer is to blame if there is a mismatch between its reading and the temperature. So there is a difference in direction of fit in this example. However, contrary to Pritchard's claim (1), it is not true that the agent's belief, after consulting the thermostat display (which is taken to be a thermometer display), has the wrong direction of fit. If beliefs are distinguished from other attitudes in terms of their direction of fit, a belief, in virtue of the kind of attitude it is, simply could not have the wrong direction of fit.

What Pritchard finds problematic in his Temp case is, I suppose, that the agent's belief about the temperature is, in the context of the world he imagines, an indirectly self-

---

the wrong direction of fit with the facts'; (2) 'We can bring this point out more clearly by considering how Temp's belief satisfied the safety principle'. 'And, for that matter, the sensitivity principle as well'; (3) Temp cannot know the temperature in this room by consulting a broken thermometer'; (4) … the underlying point demonstrates by this example is that no modal principle of the sort required to eliminate knowledge undermining luck will be able to specify the kind of direction of fit that is required for a belief to satisfy the ability intuition'.

fulfilling belief. Even directly self-fulfilling beliefs, however, do not have the wrong direction of fit. When I think I exist, my thinking that I exist makes it true that I exist. There is nothing *per se* problematic about such beliefs. We all know perfectly well that we exist.

Pritchard's example mirrors one that Roush (2005, 122) has used in her discussion of safety. Again, the world is set up so that an agent's beliefs are made true, whatever they are, by, in Roush's case, a 'fairy godmother'. I prefer Goldman's (1979) more traditional formulation of the same case in terms of a benevolent demon. Roush and Pritchard claim that an agent who, in the world of their examples, has a true temperature belief, does not know what the temperature is. They do little to explain their intuition. Pritchard (2012, 260) says that one '… cannot know the temperature in a room by consulting a broken thermometer'. This is generally true, but the device which informs the agent's beliefs in this case is not just any old broken thermometer. Due to the immediate enactment of the 'thermometer's' temperature commands, it does provide accurate temperature readings that, if we want to have modal reassurance, have the required modal stability.[36] Roush (2005, 122) says that the fact that these beliefs are self-fulfilling '…seems a happy scenario for S (although cf. King

[36] In the context of this paper, it is of interest how Pritchard (2012, 260f) explains why this agent's belief satisfies safety. 'This is ensured by the fact that the manner in which Temp is forming his beliefs, such that success is guaranteed, means that it can hardly be the case that he could have formed a false belief.' In this passage, the main lessons of this paper have been accepted. What is important is to be safe (and not to safely believe that p), and this is ensured by being sensitive to the way the world is.

Midas), but it is surely not knowledge'. I agree that there is something amiss when we consider the agent's epistemic situation. The agent misunderstands what's going on in this situation. If we rectified this, i.e. if we assumed that the agent knew that he is consulting a perfectly enacted command device, the agent would, contrary to Pritchard's claim (3), know what the temperature is (at least in this variation of the example).

In order to look for an appropriate place for a direction of fit worry, let me pursue Roush's suggestion that the world of the benevolent demon, who ensures the truth of one's beliefs, whatever they are, would be a happy world for the imagined agent. Judging from my own case, I believe quite a few things will happen which I don't want to happen. I'd rather wish no demon would come along to ensure their truth. One might still think, wrongly in my mind, that the benevolent demon provided an opportunity for happiness. Note that it will be hard to suddenly stop believing all those discomforting truths we believed all along and start believing that everything will turn out wonderful. Suppose the demon clarifies his offer. He says that everything I believe will be true for all those beliefs that I form on this very basis, i.e. on the basis that the demon will make them true. It would be very hard to take advantage of the demon's offer. How can I form a belief on the basis that whatever I come up with is going to be true? Unconstrained by the idea that what I come up with is supposed to match the world, I couldn't come up with anything we could call a belief. I could wish for whatever, and a demon who promises to make all my wishes come true would,

indeed, be welcome. I said that self-fulfilling beliefs are not problematic as beliefs. They do not have the wrong direction of fit. However, when their capacity for self-fulfilment is supposed to become the basis on which one is asked to believe, one cannot form any beliefs because beliefs have a mind-to-world direction of fit.[37]

My objection to safety, as traditionally conceived, had nothing to do with self-fulfilling beliefs. It has to do with the nature of our concern for truth. Truth itself has no direction; when an interest in truth is appropriate, we want to have matching beliefs by holding the world fixed and trying to create a match between an aspect of the world we are interested in and what we take this aspect to be. A concern for safety is not in line with the direction our epistemic concern for truth exhibits. Independently of thinking about epistemic interests, there is no legitimate direction-of-fit worry.[38]

In this section we have seen that philosophers slide between the idea of being safe from error and the notion of safely believing. Advocates of traditional safety endorse, on occasion, the view defended here: we are safe from error by being sensitive to relevant distinctions. And we have seen that the methodology we have employed, namely to think about epistemological notions in terms of our epistemic interests has

[37] For a further discussion of self-fulfilling beliefs see Piller (2016, 311-313). Pritchard's fourth claim, namely that no modal account can specify the right kind of direction of fit required for knowledge that is due to an agent's competence, has been put in doubt in section 2.2, where I argue that abilities built on appropriate sensitivity conditionals make us (reasonably) safe from error.

[38] In this section I went along with the idea that beliefs are identified by their direction of fit. No such commitment underlies my argument against safety. All I needed was the idea that an interest in truth, however it relates to the nature of believing, is an asymmetric matching concern.

an advantage: direction-of-fit worries are worries about whether epistemic notions fit into a framework of concerns that respect the asymmetric nature of our concern for truth.

4. Concluding Remarks

Sosa has introduced the safety conditional by reversing the direction of the sensitivity conditional. As only sensitivity conditionals strengthen our aim of achieving truth, the introduction of a subjunctive with the opposite direction has led to problems. These problems went unnoticed because there is a plausible idea in the vicinity: the idea of being safe. Its importance reaches far beyond epistemology into all corners of our lives. Being understandably attracted by this idea, epistemologists have neglected the difference between being safe and safely believing something. This neglect is, I've suggested, a remnant of Gettier's influence on the subject. Wanting to find a property of a belief specified by its content which could render this belief into knowledge is part of Gettier's legacy. This focus on properties of beliefs with specified content was, I have argued, a mistake. It is *us* who want to be safe from error and *we* can achieve such safety by being sensitive.[39]

[39] I have presented the main ideas of this paper at a workshop in Cologne in 2013. Sandy Goldberg, Thomas Grundmann and Ernest Sosa provided valuable feedback. I also gave this paper in Madrid in 2014 and at the universities of Haifa, Umea and Stockholm in 2017. Its last presentation was in London at the Barnes Philosophy Club in 2018. At all these occasions I've received generous comments and I am grateful to all these audiences. I had useful and encouraging discussions with Andrew Ward, David Efird, Dorothea Debus, and Tom Stoneham. Thomas Baldwin remained unconvinced. Timothy Williamson has kindly sent me his comments, as did Levi Spectre, Arnon Keren, David Sosa, Jim Pryor, Olle Risberg, and Frank Hofmann. I couldn't convince everyone. But I did convince Hannah Piller, at least she said I did.

Bibliography

Adams, F. and Clarke, M., 2005. Resurrecting the Tracking Theories. *Australasian Journal of Philosophy*, *83*(2): 207-221.

Anscombe, G.E.M. (1957). *Intention*. Harvard University Press.

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford University Press.

Blome-Tillmann, Michael (2009). Contextualism, Safety and Epistemic Relevance. *Philosophical Studies* 143 (3): 383-394.

Bogardus, Tomas (2013). Knowledge under Threat. *Philosophy and Phenomenological Research* 86 (1): 289-313.

Comesaña, Juan (2005). Unsafe knowledge. *Synthese* 146 (3): 395 - 404.

Frankfurt, Harry G. (1969). Alternate possibilities and moral responsibility. *Journal of Philosophy* 66 (3): 829-39.

Goldman, Alvin (1979). What is Justified Belief? *Justification and Knowledge*. Ed by G Pappas, Boston: D. Reidel: 1-25.

Greco, John (2007). Worries about Pritchard's Safety. *Synthese* 156: 299-302

Hawthorne, John (2014). The Case for Closure. *Contemporary Debates in Epistemology*. Ed. by M Steup, J Turri & E Sosa. Oxford: Blackwell: 27-40.

Hieronymi, Pamela (2005). The Wrong Kind of Reason. *The Journal of*

Philosophy, *102*(9): 437-457.

Humberstone, I. L. (1988). Some Epistemic Capacities. *Dialectica* 42 (3): 183-200.

Humberstone, I. L. (1992). Direction of Fit. *Mind* 101 (401): 59-83.

Kelp, Christoph (2009). Knowledeg and Safety. *Journal of Philosophical Research* 34.

McEvoy, Mark (2009). The Lottery Puzzle and Pritchard's Safety Analysis of Knowledge. *Journal of Philosophical Research* 34: 7-20.

McNicol, Don (1972). A Primer of Signal Detection Theory. London: Allen and Unwin.

Neta, Ram & Rohrbaugh, Guy (2004). Luminosity and the safety of knowledge. *Pacific Philosophical Quarterly* 85 (4): 396–406.

Nozick, Robert (1981). *Philosophical Explanations*. Oxford: Clarendon Press.

Piller, Christian (2009). Desiring the Truth and Nothing but the Truth. *Nous* 43, 2009: 193-213.

Piller, Christian (2015), Practical Philosophy and the Gettier Problem – Is Virtue Epistemology on the Right Track? *Philosophical Studies* 172: 73-91.

Piller, Christian (2015), How to Overstretch the Ethics-Epistemology Analogy: Berker's Critique of Epistemic Consequentialism. In: M Grajner (ed), *Epistemic Reasons, Epistemic Norms, and Epistemic Goals*. de Gruyter: 305-319.

Piller, Christian (2016). Evidentialism, Transparency, and Commitments. *Philosophical*

Issues 26(1): 332-350.

Pritchard, Duncan (2008). Sensitivity, Safety, and Anti-luck Epistemology. In John Greco (ed.). *The Oxford Handbook of Skepticism*. Oxford University Press: 438-455.

Pritchard, Duncan (2009). Safety-Based Epistemology: Whither Now? *Journal of Philosophical Research* 34: 33-45.

Pritchard, Duncan (2012). Anti-Luck Virtue Epistemology. *Journal of Philosophy* 109 (3): 247-279.

Roush, Sherrilyn (2005). *Tracking Truth: Knowledge, Evidence, and Science*. Oxford University Press.

Sharon, Assaf & Spectre, Levi (2017). Evidence and the Openness of Knowledge. *Philosophical Studies* 174: 1001-1037.

Sosa, Ernest (1999). How to Defeat Opposition to Moore. *Noûs*, *33*: 141-153.

Sosa, Ernest (2001). For the Love of Truth. *Virtue epistemology: Essays on Epistemic Virtue and Responsibility*: 49-62.

Sosa, Ernest (2003). The Place of Truth in Epistemology. In: M DePaul and L Zagzebski (eds.) *Intellectual virtue: Perspectives from Ethics and Epistemology*. Oxford University Press: 155-79.

Sosa, Ernest (2004). Replies. In: John Greco (ed.) (2004). *Ernest Sosa and His Critics*. Malden MA: Blackwell Publishing: 276-282.

Sosa, Ernest (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I*. OUP Oxford.

Vogel, Jonathan (2007). Subjunctivitis. *Philosophical Studies* 134 (1): 73 - 88.

Williams, Bernard (1978). *Descartes: the Project of Pure Enquiry*. New York: Penguin.

Williamson, Timothy (1996). Cognitive homelessness. *Journal of Philosophy* 93 (11): 554-573.

Williamson, Timothy (2000). *Knowledge and its Limits*. Oxford University Press.

Williamson, Timothy (2009a). Replies to Critics. In: Patrick Greenough, Duncan Pritchard & Timothy Williamson (eds.) *Williamson on Knowledge*. Oxford University Press, 280-384.

Williamson, Timothy (2009b). Probability and Danger. *The Amherst Lecture in Philosophy* 4 (2009): 1-35.